Published online 2 November 2010 *Nucleic Acids Research, 2011, Vol. 39, Database issue* ***D309–D319***
doi:10.1093/nar/gkq1009

GPCRDB: information system for G protein-coupled receptors

Bas Vroling[1], Marijn Sanders[2], Coos Baakman[1], Annika Borrmann[1], Stefan Verhoeven[2], Jan Klomp[2], Laerte Oliveira[3], Jacob de Vlieg[1,2] and Gert Vriend[1,*]

[1]CMBI, NCMLS, Radboud University Nijmegen Medical Centre, Geert Grooteplein Zuid 26-28, 6525 GA Nijmegen, [2]Department of Molecular Design and Informatics, MSD, Molenstraat 110, 5340 BH, Oss, The Netherlands and [3]Department of Biophysics, Escola Paulista de Medicina, Federal University of São Paulo, São Paulo 04023-062, Brazil

Received September 14, 2010; Accepted October 7, 2010

ABSTRACT

The GPCRDB is a Molecular Class-Specific Information System (MCSIS) that collects, combines, validates and disseminates large amounts of heterogeneous data on G protein-coupled receptors (GPCRs). The GPCRDB contains experimental data on sequences, ligand-binding constants, mutations and oligomers, as well as many different types of computationally derived data such as multiple sequence alignments and homology models. The GPCRDB provides access to the data via a number of different access methods. It offers visualization and analysis tools, and a number of query systems. The data is updated automatically on a monthly basis. The GPCRDB can be found online at http://www.gpcr.org/7tm/.

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute a large family of cell surface receptors. They regulate a wide range of cellular processes, including the senses of taste, smell and vision, and control a myriad of intracellular signaling systems in response to external stimuli. GPCRs are a major target for the pharmaceutical industry as is reflected by the fact that more than a quarter of all FDA approved drugs act on a GPCR (1). GPCRs are arguably one of the most-researched classes of proteins, but despite intensive academic and industrial research efforts over the past three decades, little is known about the structural basis of GPCR function. From about 350 genes that code for the non-olfactorial receptors in the human species (2), only about 30 are truly validated therapeutic targets (3), indicating this family's immense potential for future drug development. The fact that GPCRs can form homo-oligomeric and hetero-oligomeric complexes (4) has created a lot of new challenges and opportunities in the rational drug design process. In addition, a number of high-resolution crystal structures recently became available, providing new insights in receptor structure and function and giving the GPCR field a big stimulus.

Researchers who focus on one particular protein or a class of proteins are confronted with the fact that both the number and the size of databases are expanding at an ever-increasing pace. Although many databases like PDB (5), UniProtKB (6), KEGG (7), EMBL (8), GenBank (9), etc. are invaluable for their research, for the average wet-lab scientist these databases are less suitable for gathering, integrating, and updating different types of data in an easy and efficient manner. Studies that involve carrying over information from one protein to the other seem simple at a first glance, however, the amount of data that needs to be collected from heterogeneous sources, converted to syntactic and semantic homogeneity, validated, curated, stored and indexed, is enormous.

The GPCRDB is a data source that holds a large amount of heterogeneous data in a well-organized and easily accessible form. This data is validated, internally consistent, and updated regularly. In addition to being a one-stop GPCR resource, the data in the GPCRDB facilitates inferring new information using a wide spectrum of bioinformatics techniques. The GPCRDB is a paradigm for MCSIS technology (10,11).

NEW FEATURES

The previous release of the GPCRDB (12) was almost entirely a static website, neither offering much dynamic content, nor possibilities for complex interactions or the use of computational tools. We addressed this problem by rewriting the entire system. The use of new

*To whom correspondence should be addressed. Tel: +31 24 3619390; Fax: +31 24 3619395; Email: vriend@cmbi.ru.nl tools and modern-day e-Science technologies has resulted in improved flexibility and greater user-friendliness. We have updated the methods for harvesting GPCR sequences, expanded the number of data types available, and added new tools and services to the GPCRDB. Nearly all of the functionality that is offered through the web interface is also available in the form of web services. This allows for the easy integration of the GPCRDB in custom built tools and scripts or in workflow management tools such as Taverna (13) and Pipeline Pilot (http://www.accelrys.com/products/pipeline-pilot/). All pages now offer extensive context-sensitive help functionality, explaining what kinds of data are displayed and how to use the available interactive functionalities such as searching and computational tools.

DATA CONTENT

The contents of the GPCRDB can be categorized in three classes: primary, secondary and tertiary data. Sequence data, ligand-binding constants, mutant information, structural data and oligomer interactions make up the experimentally determined primary data. Data types such as multiple sequence alignments (MSAs), homology models, correlation patterns and entropy-variability data are inferred from these primary data, and fall in the category of secondary, or computationally derived data. Curator provided interpretations and other user help facilities make up the tertiary data category. Table 1 shows a few vital statistics about the volume of the data content of the GPCRDB.

PRIMARY DATA

Sequences

GPCR sequences are extracted from NCBI's NR database, which is a non-redundant protein sequence database with entries from a set of sequence repositories that include GenBank CDS translations, UniProtKB and PDB. GPCR sequences are selected by classifying them against a database of Hidden Markov Models (HMMs). For each of the protein families in the GPCRDB a HMM is available. These HMMs are created from MSA of the previous GPCRDB release. HMM files are created with the HMMER software package (http://hmmer.wustl.edu/).

As a first step in harvesting GPCR sequences we perform a BLAST search with all the HMM consensus sequences against the NR database. By using a very relaxed cut-off value we collect a large number of resulting hits, including many false positives. This step is necessary to limit the amount of sequences that will be used for the actual classification while ensuring a minimal false-negative rate. This step reduced the search space (for the September 2010 GPCRDB release) from about 11 million sequences to about 80 000. These hits are scored against the collection of HMMs to place them in the correct family or discard them as not being a GPCR sequence. Additional filter steps are applied such as filtering out sequence fragments and sequences that contain ambiguous amino acid characters, resulting in a final set of about thirty thousand sequences. The corresponding database entries of the selected sequences are retrieved with MRS (14) and additional data such as gene names and species information is extracted and stored. The GPCRDB holds for each sequence one principal access page. Figure 1 shows an example of such a page.

Table 1. Statistics for the September 2010 release of the GPCRDB

Sequences	27 045
Families (and MSA)	1270
Mutations	7703
Ligand-binding data	12 086
Protein structures	195
Homology models	22 616
Residues	11 290 993
Species	1521
Oligomers	115

The protein detail page (Figure 1) contains a panel that visualizes sequence annotations such as helix boundaries, cysteine bridges and glycosylation sites. These annotations are loaded in real time using the DAS distributed annotation system (15,16) and are visualized by Dasty2 (17), resulting in always up-to-date annotations. We use the UniProt DAS server to retrieve sequence annotations.

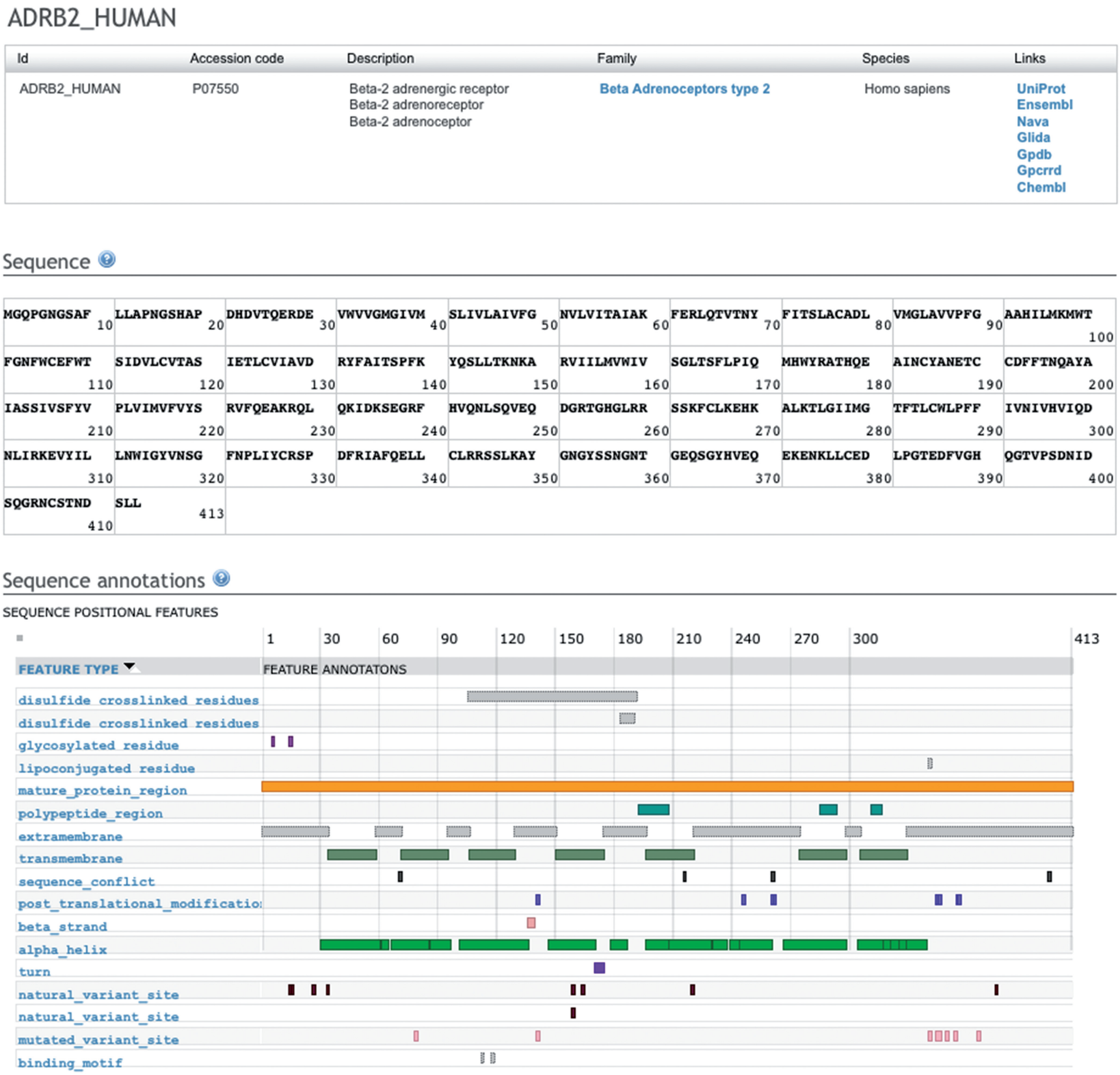
Ligand-binding data

Ligand-binding constants are available for a large number of GPCRs and are obtained from various sources. For each GPCR we provide links, if possible, to the ChEMBL (18) and GLIDA (19) databases. In addition, ligand-binding information that is obtained from collections from Seeman (20) and Organon (21) is available. Since ligand-binding data is very hard to obtain from the literature we encourage academic and industrial researchers to submit their ligand-binding data to the GPCRDB in order to make this data accessible to the scientific community.

Mutations

The GPCRDB contains a large number of well-annotated mutations obtained from different sources. We have two sets of mutations that were manually extracted from literature. Mutant data from the tinyGRAP database (22) contains references to scientific literature describing point mutations as well as insertions, deletions and chimeric receptors. A collection of in-house manually extracted mutant data contains a few thousand point mutations and the effects of those mutations on the function of the receptor. We have extracted sentences from the papers that qualitatively describe the effects of these mutation and we have extracted quantitative data such as effects on ligand binding, expression, activation or constitutive activity.

In addition to the two manually curated data sets we also have a large body of mutations that were extracted from the literature by the software package MuteXt (23). A sentence describing the effects of the mutations is available for all mutations extracted by MuteXt.



ADRB2_HUMAN

Id	Accession code	Description	Family	Species	Links
ADRB2_HUMAN	P07550	Beta-2 adrenergic receptor Beta-2 adrenoreceptor Beta-2 adrenoceptor	Beta Adrenoceptors type 2	Homo sapiens	UniProt Ensembl Nava Glida Gpdb Gpcrrd Chembl

Sequence

MGQPGNGSAF 10	LLAPNGSHAP 20	DHDVTQERDE 30	VWVVGMGIVM 40	SLIVLAIVFG 50	NVLVITAIAK 60	FERLQTVTNY 70	FITSLACADL 80	VMGLAVVPFG 90	AAHILMKMWT 100
FGNFWCEFWT 110	SIDVLCVTAS 120	IETLCVIAVD 130	RYFAITSPFK 140	YQSLLTKNKA 150	RVIILMVWIV 160	SGLTSFLPIQ 170	MHWYRATHQE 180	AINCYANETC 190	CDFFTNQAYA 200
IASSIVSFYV 210	PLVIMVFVYS 220	RVFQEAKRQL 230	QKIDKSEGRF 240	HVQNLSQVEQ 250	DGRTGHGLRR 260	SSKFCLKEHK 270	ALKTLGIIMG 280	TFTLCWLPFF 290	IVNIVHVIQD 300
NLIRKEVYIL 310	LNWIGYVNSG 320	FNPLIYCRSP 330	DFRIAFQELL 340	CLRRSSLKAY 350	GNGYSSNGNT 360	GEQSGYHVEQ 370	EKENKLLCED 380	LPGTEDFVGH 390	QGTVPSDNID 400
SQGRNCSTND 410	SLL 413								

Sequence annotations

SEQUENCE POSITIONAL FEATURES

Figure 1. Screenshot of the principal protein sequence page of the human beta-2 adrenoceptor. The top table contains details about the protein record and hyperlinks to the protein family browsing page and other data sources that contain information about this protein. The middle table holds the sequence in which each amino acid is linked to its own residue page. The bottom table holds annotations that are obtained in real time using the DAS (15) system.

Structures

Structures are obtained from the PDB. Links to structures that were re-refined in the PDB_REDO project (24) are included. We provide manually 'cleaned' monomers of the major GPCR PDB files that have been prepared for easy casual use by the life sciences community.

Oligomers

GPCR oligomerization has been an area of interest and controversy for many years. Recently there has been increasing evidence that both homo-dimers and hetero-dimers play a crucial role in GPCR signaling (25–27). The GPCR-OKB (28) is a database that stores manually extracted computational and experimental information about GPCR oligomerization. Lists of protomers, experimental details and, where available, inferred oligomer interaction sites are available for all oligomers. This data has been fully integrated in the GPCRDB, making the information about both GPCR protomers and oligomers readily available.

SECONDARY DATA

Multiple sequence alignments

MSAs are available for all families. MSAs are generated with WHAT IF (29) for all GPCR sub-families using hand-optimized sub-family specific profiles. Position-specific annotations such as secondary structure information and generalized residue numbers are stored in the profiles and are incorporated in the alignments. The general residue numbers are relative to the arbitrarily selected numbers for very conserved residues and motifs such as the well-known E/DRY and NPXXY motifs. Using a profile to align a GPCR sub-family allows for the mapping of the general residue numbers on the sequences that are being aligned. The result is that the residues in the TM domains, helix VIII and sections of the loops are labeled with a general residue number.

For creating alignments of parent GPCR families we make use of these general residue numbers. For all the GPCRs that are being aligned we select all the general residue positions that the sub-families have in common and create the alignment by listing, for each sequence, the residues at the selected positions. GPCR parent family alignments are thus not built using standard alignment algorithms but are created by selecting residues that are likely to share the same position in the 3D structure.

Homology models

Despite the recent publication of a number of GPCR structures the amount of structural information on GPCRs is still very limited. We have built structure models of all Class A receptors to at least partially fill this gap. Information extracted from the profile-based MSA is used to generate high-quality sequence–structure alignments between GPCR sequences and a number of experimentally determined structures. Based on these alignments homology models of all GPCRs have been automatically created and will be automatically updated as more sequences become available. Template structures are selected based on sequence identity and a number of structure quality criteria. The homology models are created with WHAT IF and YASARA (30). Models will be automatically replaced when new structures become available that are better templates, with better being defined as either being solved with higher resolution data, or as having a higher percentage sequence identity.

Correlated mutation analyses

Correlated mutation analysis (CMA) is a technique that can find pairs of residues that remain conserved or mutate in tandem during evolution. Residues that show correlated behaviour in MSA are likely to be functionally related, and networks of those correlating residues indicate functional groups (31). The rationale behind this analysis is that when a mutation occurs at a functionally important site, the protein either becomes functionally impaired or may acquire its original or a different function due to compensatory mutations at one or more other positions. Correlation scores are available in a number of different formats for all GPCR (sub)families.

Entropy and variability data

The amount of entropy and variability that is observed for a certain position in multiple sequence alignment tells something about the types of pressures exerted on that position during evolution (32–34). Entropy-variability data is available for each multiple sequence alignment in the GPCRDB. We offer this data in tabular from, entropy-variability plots (35), and more advanced sub-family specific two-entropy plots (manuscript in preparation) based on the original method described by Ye *et al.* (36) (Figure 2).

TERTIARY DATA

Residue annotations

Residues in the GPCRDB are labelled with the original Oliveira *et al.* (37) numbering scheme as well as with residue numbers from the more recent Ballesteros–Weinstein scheme. Use of these general residue numbers allows for easy transfer of information between proteins. General residue numbers are available for all residues within conserved structure elements. These include all transmembrane helices, helix VIII and a few sections in the loops. For each residue with a general residue number a short description of its properties and interactions is available. These descriptions are based on a manual analysis of the currently available crystal structures.

Cytoscape networks

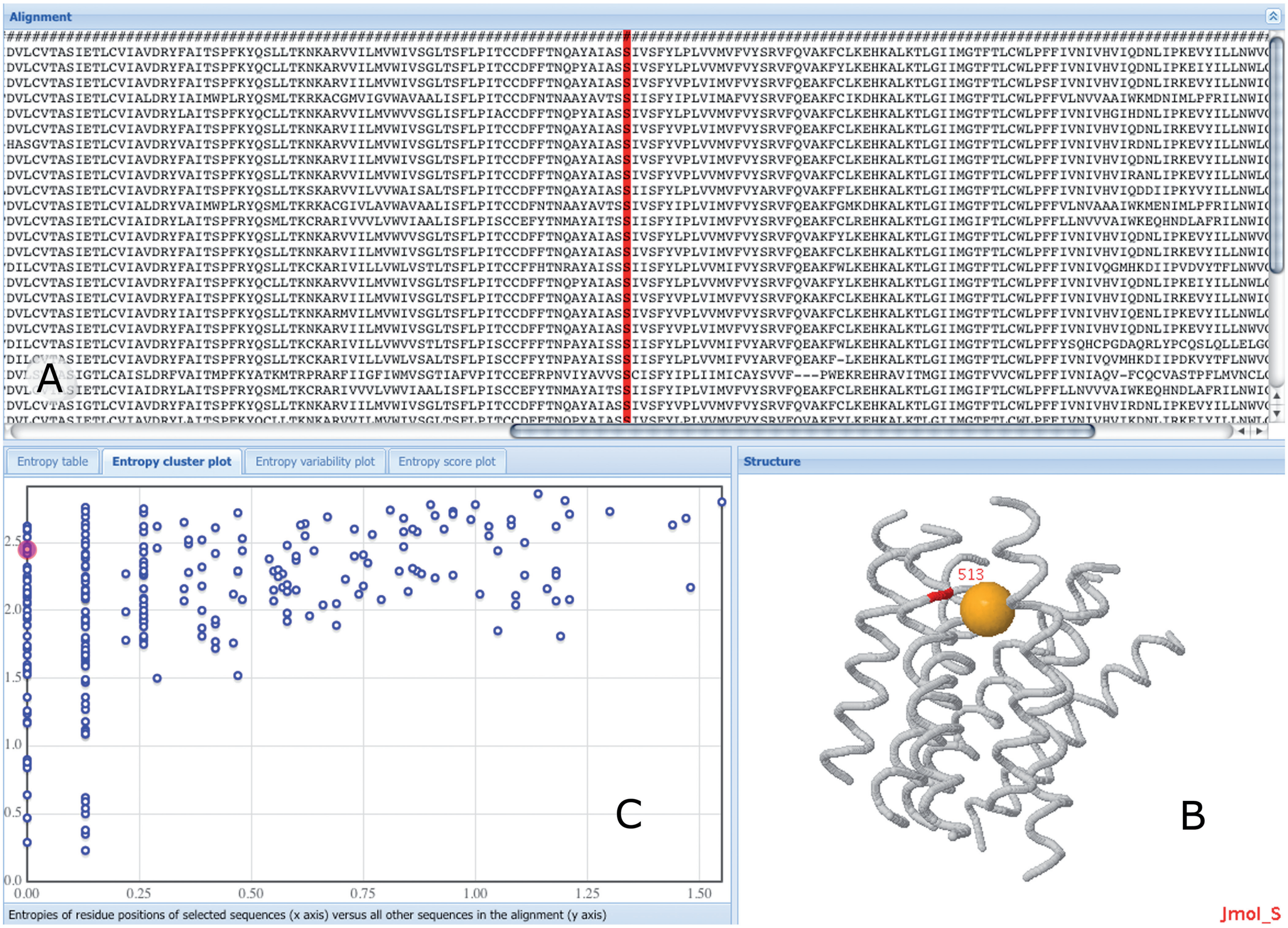
The GPCRDB provides cytoscape (38) network files for all GPCR families. These network files contain the proteins of a family with distances calculated from the family alignments. For all proteins the protein family information, species names and the amino acid types for all the residues annotated with a general residue number are available as attributes. This allows for complex analyses, such as coloring proteins by amino acids at a certain residue position to compare i.e. species or sub-type specific differences.

Mutation predictions

For all positions for which a general residue number is available we have investigated the most likely effects of mutations at these positions. Short fragments of text have been created that explain for each of these positions the likely effects of the mutation on structural and functional levels. References to key papers in which experimental evidence for these effects is available are included in the fragments. Examples of such effects are the loss of ligand-binding affinity when mutating a residue in the ligand-binding pocket, the loss of G protein binding when mutating residues at the G protein-binding interface, and increased constitutive activity when residues are mutated a the interface between helix III and VI.

Workflows

We have created a number of Taverna workflows that use the web services of the GPCRDB as a starting point for users who want to programmatically access the GPCRDB.



```
###################################################################################################################################################
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYLPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIPKEVYILLNWVC
DVLCVTASIETLCVIAVDRYFAITSPFKYQCLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQPYAIASSIVSFYLPLVVMVFVYSRVFQVAKFYLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIPKEIYILLNWLC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKFCLKEHKALKTLGIIMGTFTLCWLPSFIVNIVHVIQDNLIRKEVYILLNWIC
DVLCVTASIETLCVIALDRYIAIMWPLRYQSMLTKRKACGMVIGVWAVAALISFLPITCCDFNTNAAYAVTSSIISFYIPLVIMAFVYSRVFQEAKFCIKDHKALKTLGIIMGTFTLCWLPFFVLNVVAAIWKMDNIMLPFRILNWIC
DVLCVTASIETLCVIAVDRYLAITSPFKYQCLLTKNKARVVILMVWVVSGLISFLPIACCDFFTNQPYAIASSIVSFYLPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHGIHDNLIPKEVYILLNWVC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIC
HASGVTASIETLCVIAVDRYVAITSPFKYQSLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRDNLIPKEVYILLNWLC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIC
DVLCVTASIETLCVIAVDRYVAITSPFKYQSLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRANLIPKEVYILLNWLC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKSKARVVILVVWAISALTSFLPITCCDFFTNQAYAIASSIISFYLPLVVMVFVYARVFQVAKFFLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDDIIPKYVYILLNWLC
DVLCVTASIETLCVIALDRYVAIMWPLRYQSMLTKRKACGIVLAVWAVAALISFLPITCCDFNTNAAYAVTSSIVSFYIPLVIMVFVYSRVFQEAKFGMKDHKALKTLGIIMGTFTLCWLPFFVLNVAAAIWKMENIMLPFRILNWIC
DVLCVTASIETLCVIAIDRYLAITSPFRYQSMLTKCRARIVVVLVWVIAALISFLPISCCEFYTNMAYAITSSIISFYIPLVIMVFVYSRVFQEAKFCLREHKALKTLGIIMGIFTLCWLPFFLLNVVVAIWKEQHNDLAFRILNWIC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVVILMVWVVSGLTSFLPITCCDFFTNQAYAIASSIVSFYLPLVVMVFVYSRVFQVAKFYLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIPKEVYILLNWVC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIC
DILCVTASIETLCVIAVDRYFAITSPFRYQSLLTKCKARIVILLVWLVSTLTSFLPITCCFFHTNRAYAISSSIISFYLPLVVMIFVYSRVFQEAKFWLKEHKALKTLGIIMGTFTLCWLPFFIVNIVQGMHKDIIPVDVYTFLNWVC
DVLCVTASIETLCVIAVDRYFAITSPFKYQCLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQPYAIASSIVSFYLPLVVMVFVYSRVFQVAKFYLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIPKEIYILLNWLC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQKAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIC
DVLCVTASIETLCVIAVDRYIAITSPFKYQSLLTKNKARMVILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQENLIPKEVYILLNWLC
DVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIPKEVYILLNWVC
DILCVTASIETLCVIAVDRYFAITSPFRYQSLLTKCKARIVILLVWVVSTLTSFLPISCCFFFTNPAYAISSSIISFYLPLVVMIFVYARVFQEAKFWLKEHKALKTLGIIMGTFTLCWLPFFYSQHCPGDAQRLYPCQSLQLLELGC
DILCVTASIETLCVIAVDRYFAITSPFRYQSLLTKCKARIVILLVWLVSALTSFLPISCCFFYTNPAYAISSSIISFYLPLVVMIFVYARVFQEAKF-LKEHKALKTLGIIMGTFTLCWLPFFIVNIVQVMHKDIIPDKVYTFLNWVC
DVLSVASIGTLCAISLDRFVAITMPFKYATKMTRPRARFIIGFIWMVSGTIAFVPITCCEFRPNVIYAVVSSCISFYIPLIIMICAYSVVF---PWEKREHRAVITMGIIMGTFVVCWLPFFIVNIAQV-FCQCVASTPFLMVNCLC
DVLCVTASIETLCVIAIDRYLAITSPFRYQSMLTKCRARIVVVLVWVIAALISFLPISCCEFYTNMAYAITSSIISFYIPLVIMVFVYSRVFQEAKFCLREHKALKTLGIIMGIFTLCWLPFFLLNVVVAIWKEQHNDLAFRILNWIC
DVLCVTASIGTLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGLTSFLPITCCDFFTNQAYAIASSIVSFYVPLVVMVFVYSRVFQVAKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRDNLIPKEVYILLNWVC
DVLCVTASIETLCVIAVDRYLAITSPFKYQCLLTKNKARVVILMVWIVSGLTSFLPITCCDFFTNQPYAIASSIVSFYLPLVVMVFVYSRVFQVAKFYLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIKDNLIRKEIYILLNWLC
```

Figure 2. Screenshot of the interactive entropy and variability page. The multiple sequence alignment is shown in (**A**). Residues are interlinked in all page elements, clicking results in highlighted selections. In (**B**), the approximate location of this position in the 3D model of the transmembrane domain of class A GPCRs is shown in red and is annotated with general residue number information. The orange ball in the structure model indicates the approximate location of the assumed binding site for low molecular weight compounds of class A GPCRs. In (**C**) the user can choose among four display modes that describe the entropy and variability of all positions in the alignment; shown is the entropy cluster variant.

Workflows are available that use the GPCRDB BLAST service, create custom-built alignments and retrieve several different data types. The workflows and documentation are available via the myExperiment web portal (39) and are tagged with 'GPCRDB'. We encourage researchers to share their own workflows via the myExperiment portal.

DATA ACCESS

The GPCRDB provides fast and easy access to all its data and information. The GPCRDB does not merely lists available data, but all data types are fully integrated. For example, mutations are accessible via the protein detail pages but can also be found at the residue level. The same holds true for oligomer interaction interfaces, where details about these interactions are available via the oligomer pages, but also via the protein detail pages of interacting members, as well as via the pages of residues that are reported in the interaction. This tight data integration makes it a very efficient resource to use. The GPCRDB's user interface allows the user to easily navigate from one data type to another and often suggests multiple routes to explore the data, thereby hopefully generating ideas and questions while the user navigates the system.

The four fundamental facilities to be provided by information systems are browsing, retrieval, query and inferencing. These four types of access have been an integral part of the GPCRDB set-up from the beginning. The total redesign of the GPCRDB that has taken place the past few years has allowed us to add novel access facilities in all these four categories.

Browsing

The main way to access the data is via a hierarchical list of GPCR families, which is based on the pharmacological classification of GPCRs (40) (Figure 3). Users can traverse the GPCR family tree and view or download the data for a selected family. Available data types include MSAs, entropy-variability analyses and lists of family members. Alignments can be viewed in multiple ways. In addition to the classic HTML view, the GPCRDB offers an interactive multiple sequence alignment viewer [JalView (41)], that can show additional information about the MSAs, supports a number of viewing and sorting options, and that can be used to generate phylogenetic trees. Residues for which mutation data is available are hyperlinked in the alignments to pages that contain more details about those mutations.

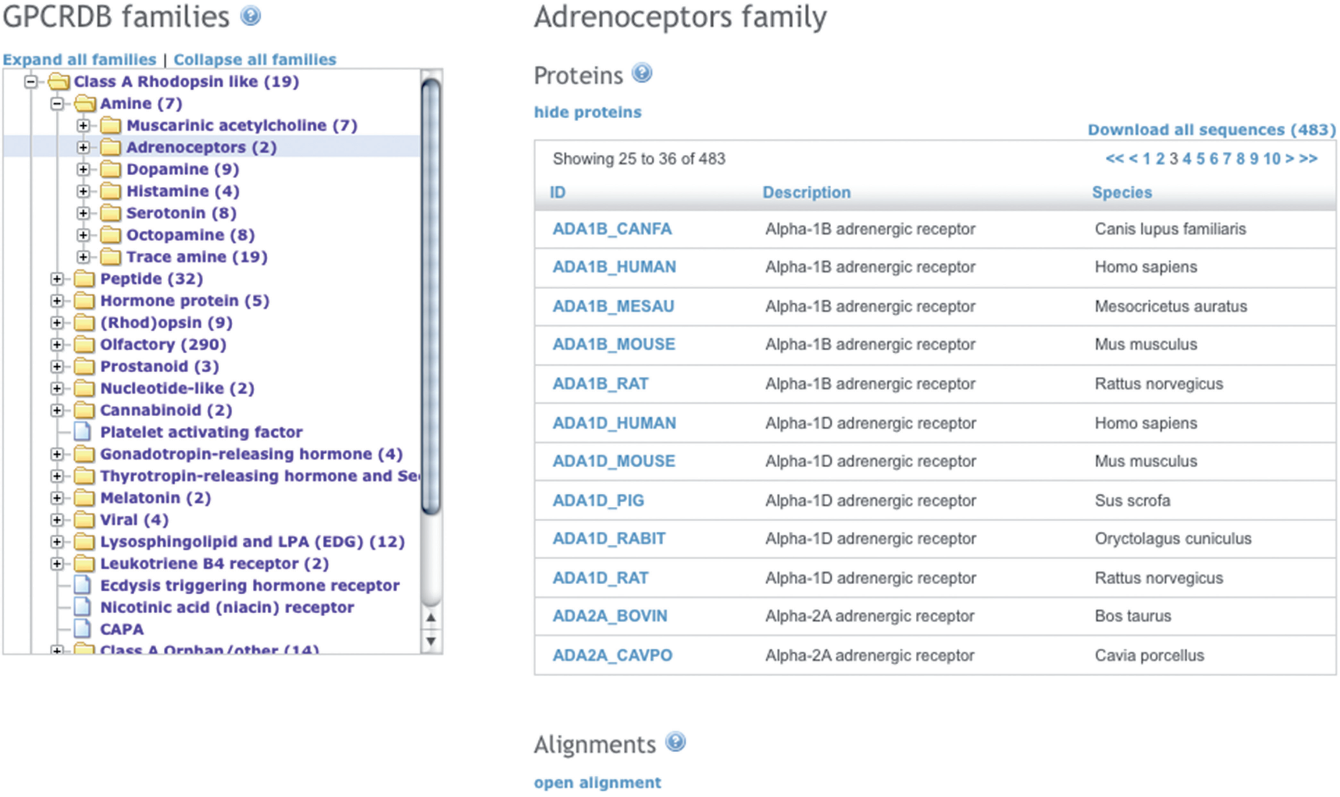
The pages that display data on individual proteins (Figure 1) contain a large amount of data and links to other data sources. Table 2 lists the remote databases that have been indexed in the GPCRDB. Some of these remote data are actually most easily queried via the GPCRDB.

On the protein detail page the sequence is displayed and each residue is hyperlinked to its individual page where additional information is listed about that specific residue; residue numbers in multiple numbering schemes, available mutation data for that specific residue and mutations at equivalent positions, and reported oligomeric interactions.

Snake plots are available for all proteins in the GPCRDB (Figure 4). Residues for which mutations are available are hyperlinked from the snake plots to pages that contain details about those mutations.

Other data types such as mutations, ligand-binding constants and information about oligomerization states are displayed when available and links are provided to pages that contain more detailed information about those data (Figure 5).

The pages with mutation details contain links to the scientific literature and if available in that literature, the qualitative and quantitative data on the effects of



GPCRDB families

Expand all families | Collapse all families

- Class A Rhodopsin like (19)
  - Amine (7)
    - Muscarinic acetylcholine (7)
    - Adrenoceptors (2)
    - Dopamine (9)
    - Histamine (4)
    - Serotonin (8)
    - Octopamine (8)
    - Trace amine (19)
  - Peptide (32)
  - Hormone protein (5)
  - (Rhod)opsin (9)
  - Olfactory (290)
  - Prostanoid (3)
  - Nucleotide-like (2)
  - Cannabinoid (2)
  - Platelet activating factor
  - Gonadotropin-releasing hormone (4)
  - Thyrotropin-releasing hormone and Se
  - Melatonin (2)
  - Viral (4)
  - Lysosphingolipid and LPA (EDG) (12)
  - Leukotriene B4 receptor (2)
  - Ecdysis triggering hormone receptor
  - Nicotinic acid (niacin) receptor
  - CAPA
  - Class A Orphan/other (14)

Adrenoceptors family

Proteins

hide proteins

Download all sequences (483)

Showing 25 to 36 of 483

<< < 1 2 3 4 5 6 7 8 9 10 > >>

ID	Description	Species
ADA1B_CANFA	Alpha-1B adrenergic receptor	Canis lupus familiaris
ADA1B_HUMAN	Alpha-1B adrenergic receptor	Homo sapiens
ADA1B_MESAU	Alpha-1B adrenergic receptor	Mesocricetus auratus
ADA1B_MOUSE	Alpha-1B adrenergic receptor	Mus musculus
ADA1B_RAT	Alpha-1B adrenergic receptor	Rattus norvegicus
ADA1D_HUMAN	Alpha-1D adrenergic receptor	Homo sapiens
ADA1D_MOUSE	Alpha-1D adrenergic receptor	Mus musculus
ADA1D_PIG	Alpha-1D adrenergic receptor	Sus scrofa
ADA1D_RABIT	Alpha-1D adrenergic receptor	Oryctolagus cuniculus
ADA1D_RAT	Alpha-1D adrenergic receptor	Rattus norvegicus
ADA2A_BOVIN	Alpha-2A adrenergic receptor	Bos taurus
ADA2A_CAVPO	Alpha-2A adrenergic receptor	Cavia porcellus

Alignments

open alignment

Figure 3. Screenshot of the GPCR family page. The GPCR family tree is shown on the left with the amine sub-family expanded. On the right-hand side the data for the selected family (adrenoceptors) is shown.

Table 2. Non-GPCRDB data facilities that can be found through the GPCRDB

Database	Type of data	Address
GPCR-OKB (GPCR oligomerization knowledge base)	Dimer information	http://data.gpcr-okb.org/gpcr-okb/
GPCRRD (GPCR restraint database)	Modeling restraints	http://zhanglab.ccmb.med.umich.edu/GPCRRD/
GLIDA (GPCR ligand database)	Ligand data	http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/
gpDB (G protein database)	G Protein data	http://biophysics.biol.uoa.gr/gpDB/
UniProt	Protein information	http://www.uniprot.org/
NAVA	Sequence Variants	http://nava.liacs.nl/
ChEMBL	Ligand data	http://www.ebi.ac.uk/chembl/
Ensembl	Genomic information and annotations	http://www.ensembl.org

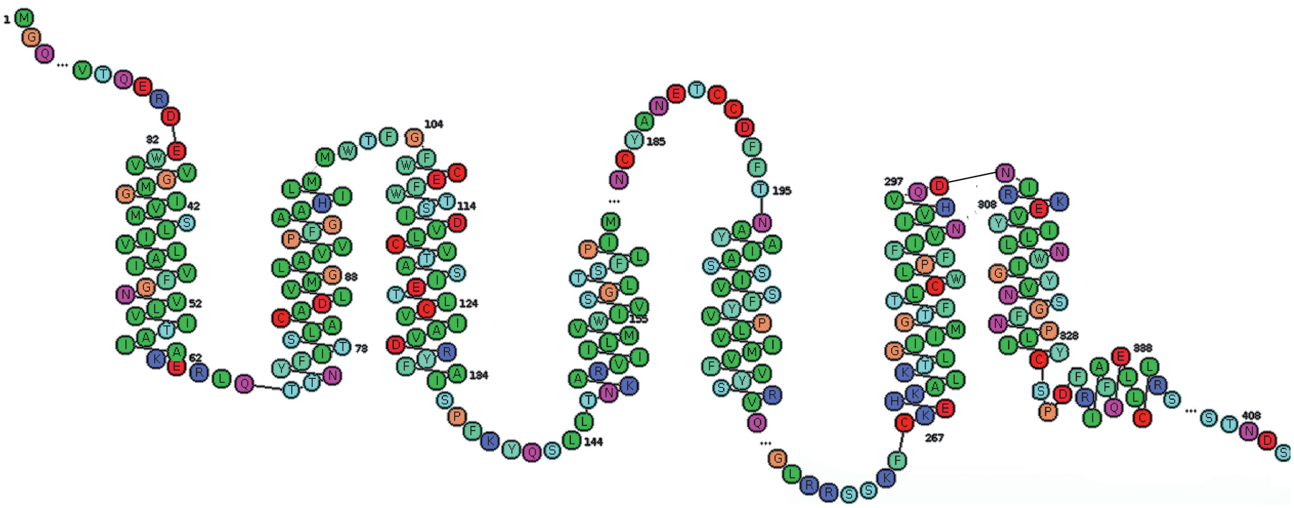
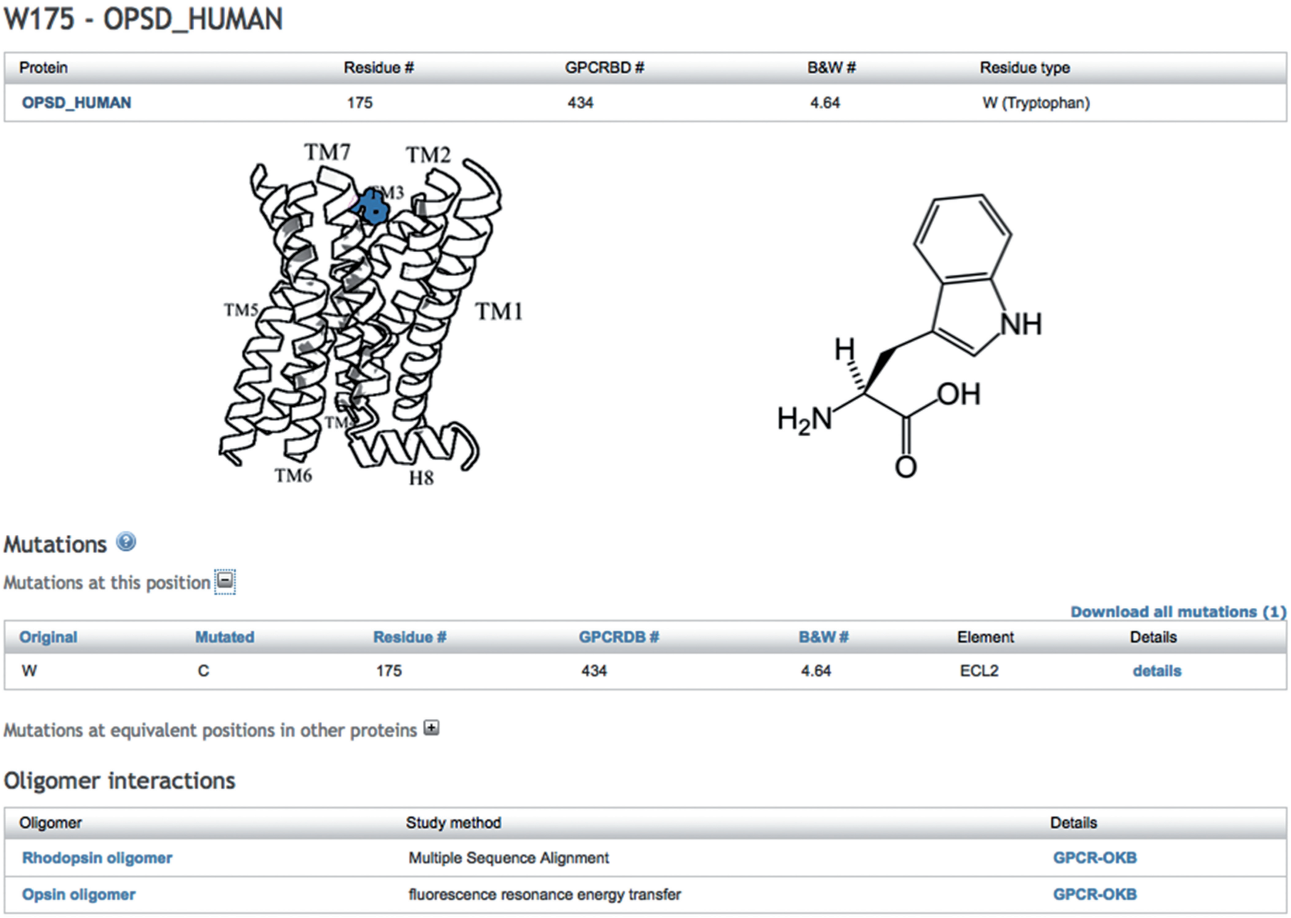


Figure 4. Snake plot of the human β2 adrenoceptor.



W175 - OPSD_HUMAN

Protein	Residue #	GPCRBD #	B&W #	Residue type
OPSD_HUMAN	175	434	4.64	W (Tryptophan)

Mutations

Mutations at this position

Download all mutations (1)

Original	Mutated	Residue #	GPCRDB #	B&W #	Element	Details
W	C	175	434	4.64	ECL2	details

Mutations at equivalent positions in other proteins

Oligomer interactions

Oligomer	Study method	Details
Rhodopsin oligomer	Multiple Sequence Alignment	GPCR-OKB
Opsin oligomer	fluorescence resonance energy transfer	GPCR-OKB

Figure 5. Screenshot of the detail page of residue W175 in OPSD_HUMAN. Residue numbers in different formats are shown, the approximate location of the residue is visually indicated and available mutations and oligomer data for this residue are listed.

the mutations is displayed. The oligomer detail page contains links to the GPCR-OKB and to the individual protomers in the GPCRDB. If certain residues are involved in the oligomer interaction, hyperlinks to individual residue pages are available.

Retrieval

Data can be retrieved via the web pages and via the web services. The web services offer very extensive retrieval possibilities, together allowing for the retrieval of all data types present in the GPCRDB. Subsets have been created for frequently requested data sets such as all human sequences and the human non-olfactory sequences. Protein family alignments can be downloaded in a number of different formats. Sequences, structures, ligand-binding data and mutations can be downloaded from the protein detail pages. After querying the GPCRDB via the web pages, query result sets can be downloaded in FASTA format. A complete copy of the GPCRDB is freely available for in-house usage by academic and industrial researchers alike.

Query

Users can query the GPCRDB via a number of different search pages. Identifiers, genes, species, descriptions and protein family names can be used to search for GPCRs (Figure 6). There are a number of filters available to limit the search results. Users can indicate whether only GPCRs should be shown for which mutations, structures, oligomers, or ligand-binding data are available.

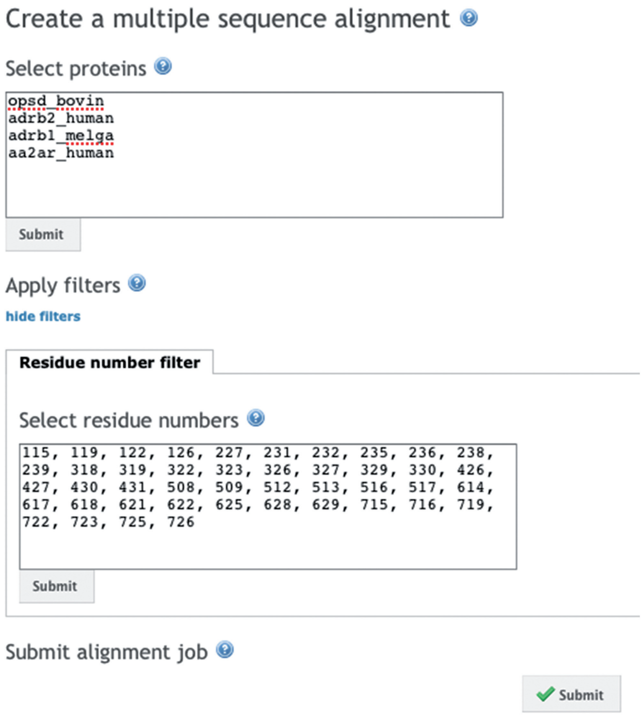
Mutations can be found via the mutant search page, where one can search by residue number (multiple numbering schemes are available) and/or residue types. The GPCRDB offers a BLAST service that allows users to BLAST their sequence against the sequences in the GPCRDB.

All search options and the BLAST services are available via the web interface and as web services. A full SQL search facility will be made available in the near future to allow for complex queries and analyses.

Inferences

The amount of available GPCR related data is too large for a human to grasp and disseminate. The GPCRDB contains a series of inference engines that determine interesting correlations between the data, while a series of software tools help the user with data reduction and abstraction.

Building alignments. The GPCRDB offers the possibility to create custom-made alignments. The alignments are created by using the procedure that is used for the parent GPCR families as discussed earlier. Users can select the proteins and residue positions that should be aligned, allowing for the creation of e.g. an alignment of all binding pocket residues for a selection of proteins (Figure 7). The custom-built alignments are available for download and users can directly interact with the alignments using JalView.



Create a multiple sequence alignment

Select proteins

```
opsd_bovin
adrb2_human
adrb1_melga
aa2ar_human
```

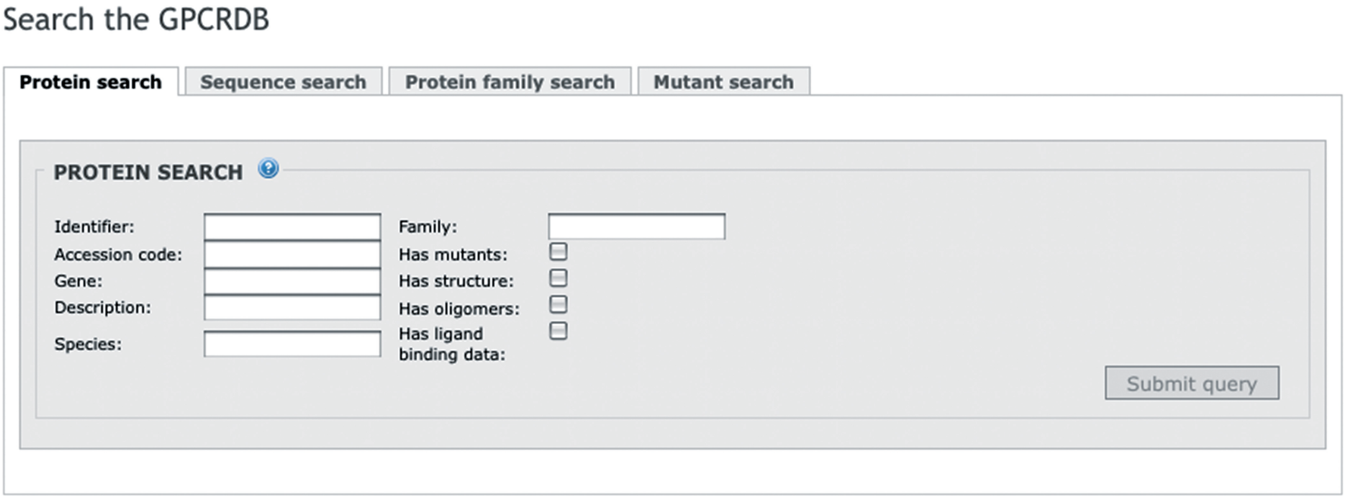
Apply filters

Select residue numbers

```
115, 119, 122, 126, 227, 231, 232, 235, 236, 238,
239, 318, 319, 322, 323, 326, 327, 329, 330, 426,
427, 430, 431, 508, 509, 512, 513, 516, 517, 614,
617, 618, 621, 622, 625, 628, 629, 715, 716, 719,
722, 723, 725, 726
```

Submit alignment job

Figure 7. A list of proteins and an optional list of residue positions can be used to generate custom alignments. In this figure we have selected a number of proteins for which crystal structures are available. The GPCR-binding pocket residue positions as proposed by Gloriam *et al.* (42) are used. The result will be an alignment of all pocket residues of the selected proteins.



Search the GPCRDB

PROTEIN SEARCH

Identifier: ____ Family: ____
Accession code: ____ Has mutants: □
Gene: ____ Has structure: □
Description: ____ Has oligomers: □
Species: ____ Has ligand binding data: □

Figure 6. The protein search page.

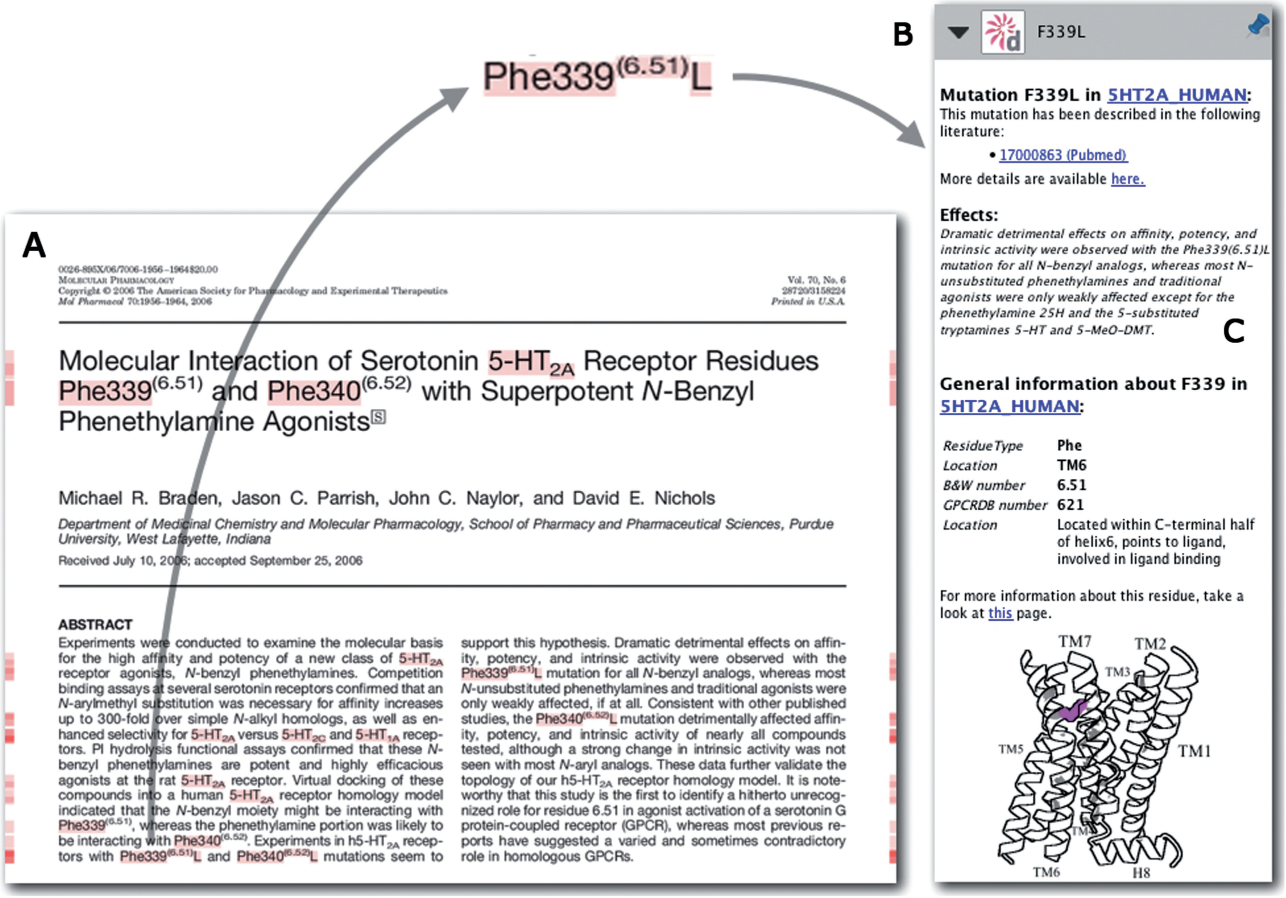
Predicting the effects of mutations. We have started to create a service where users can predict the effects of a point mutation. As for now, predictions are mainly based on human knowledge that is stored in a computer readable format. This information is combined with a number of simple analyses on a homology model of the receptor being mutated, such as looking for steric clashes and helix disruptions. Results are presented as text-fragments that explain the effects of the mutation on the structure. Care has been taken to ensure that the results are presented in a life scientist friendly manner. The text contains references to literature and is enriched with figures and animations of the mutation and its surrounding environment. In the near future more intelligence will be added to the software, such as incorporating the quantitative data from the mutations extracted from literature and ligand-binding information.

Analysis of entropy derived patterns. We offer a page where users can interactively analyze the protein family alignments (Figure 2). Plots displaying entropy and variability scores are displayed with a 3D model and a MSA. Residues are linked in the three page elements, so that clicking on a residue position in the multiple sequence alignment will highlight the residue position in the structure as well as in the entropy and variability plots. This offers researchers a very intuitive way of looking at conservation scores, even at the subfamily or receptor level, and relating those scores to the 3D structure. In combination with above mentioned accessible data site directed mutagenesis candidate selection, homology modeling and ligand-binding hypotheses generation can be performed.

Annotating scientific literature. We have developed a new interface for the GPCR data in the form of a GPCR-specific PDF reader (manuscript in preparation). This reader can annotate scientific literature on GPCRs on the fly, providing users with context sensitive data from the GPCRDB (Figure 8). This software is available upon request and will be made freely available at the day of publication of this article.

IMPLEMENTATION

The data in the GPCRDB is stored in a PostgreSQL (http://www.postgresql.org/) relational database. The web service interface is developed with the Apache CXF



0026-895X/06/7006-1956–1964$20.00
Molecular Pharmacology
Copyright © 2006 The American Society for Pharmacology and Experimental Therapeutics
Mol Pharmacol 70:1956–1964, 2006

Vol. 70, No. 6
28720/3158224
Printed in U.S.A.

Molecular Interaction of Serotonin 5-HT$_{2A}$ Receptor Residues Phe339$^{(6.51)}$ and Phe340$^{(6.52)}$ with Superpotent *N*-Benzyl Phenethylamine Agonists[S]

Michael R. Braden, Jason C. Parrish, John C. Naylor, and David E. Nichols

Department of Medicinal Chemistry and Molecular Pharmacology, School of Pharmacy and Pharmaceutical Sciences, Purdue University, West Lafayette, Indiana

Received July 10, 2006; accepted September 25, 2006

ABSTRACT

Experiments were conducted to examine the molecular basis for the high affinity and potency of a new class of 5-HT$_{2A}$ receptor agonists, *N*-benzyl phenethylamines. Competition binding assays at several serotonin receptors confirmed that an *N*-arylmethyl substitution was necessary for affinity increases up to 300-fold over simple *N*-alkyl homologs, as well as enhanced selectivity for 5-HT$_{2A}$ versus 5-HT$_{2C}$ and 5-HT$_{1A}$ receptors. PI hydrolysis functional assays confirmed that these *N*-benzyl phenethylamines are potent and highly efficacious agonists at the rat 5-HT$_{2A}$ receptor. Virtual docking of these compounds into a human 5-HT$_{2A}$ receptor homology model indicated that the *N*-benzyl moiety might be interacting with Phe339$^{(6.51)}$, whereas the phenethylamine portion was likely to be interacting with Phe340$^{(6.52)}$. Experiments in h5-HT$_{2A}$ receptors with Phe339$^{(6.51)}$L and Phe340$^{(6.52)}$L mutations seem to support this hypothesis. Dramatic detrimental effects on affinity, potency, and intrinsic activity were observed with the Phe339$^{(6.51)}$L mutation for all *N*-benzyl analogs, whereas most *N*-unsubstituted phenethylamines and traditional agonists were only weakly affected, if at all. Consistent with other published studies, the Phe340$^{(6.52)}$L mutation detrimentally affected affinity, potency, and intrinsic activity of nearly all compounds tested, although a strong change in intrinsic activity was not seen with most *N*-aryl analogs. These data further validate the topology of our h5-HT$_{2A}$ receptor homology model. It is noteworthy that this study is the first to identify a hitherto unrecognized role for residue 6.51 in agonist activation of a serotonin G protein-coupled receptor (GPCR), whereas most previous reports have suggested a varied and sometimes contradictory role in homologous GPCRs.

Mutation F339L in 5HT2A_HUMAN:
This mutation has been described in the following literature:

- 17000863 (Pubmed)

More details are available here.

Effects:
Dramatic detrimental effects on affinity, potency, and intrinsic activity were observed with the Phe339(6.51)L mutation for all N-benzyl analogs, whereas most N-unsubstituted phenethylamines and traditional agonists were only weakly affected except for the phenethylamine 25H and the 5-substituted tryptamines 5-HT and 5-MeO-DMT.

General information about F339 in 5HT2A_HUMAN:

ResidueType	Phe
Location	TM6
B&W number	6.51
GPCRDB number	621
Location	Located within C-terminal half of helix6, points to ligand, involved in ligand binding

For more information about this residue, take a look at this page.

Figure 8. An impression of the PDF reader [Utopia Documents (43), Utopia Documents-GPCRDB (in preparation)] interface to the GPCRDB data. On the left side (**A**) a scientific paper (44) is shown that is annotated by the GPCRDB. Annotations are available for all the highlighted words. On the right side (**B**) an example of such an annotation (the mutation F339L) is displayed. A short, manually extracted description of the effects of this mutation is included (**C**).

(http://cxf.apache.org/) web services framework. We offer both SOAP and REST endpoints. The web interface is built using the Apache Wicket (http://wicket.apache.org/) web application framework. The database is accessed via a Hibernate (http://www.hibernate.org) object-relational mapping layer. The server is running within Sun's Glassfish (http://glassfish.org) application server.

FUTURE DIRECTIONS

In the near future we would like to extend the interactive facilities of the GPCRDB by offering users more tools to analyze the available data. The entropy-variability analysis pages are a good example of the types of services we will be offering. In addition to our main focus of data collection and integration we would like to extend our focus towards the more challenging field of knowledge integration. The mutation effect predictor is a pilot project to explore the things we can do by combining human expertise with computational power. We are in the process of transforming the GPCRDB from mainly a one-stop resource for GPCRDB data to a place where scientists can use tools to interact with the data and make predictions.

ACKNOWLEDGEMENTS

The authors thank Bob Bywater for stimulating discussions, and Maarten Hekkelman and Tim te Beek for their support with computer science issues. Steve Pettifer, Dave Thorne and Phil McDermott are thanked for providing us with the Utopia Documents PDF reader and their support with connecting it to the GPCRDB.

FUNDING

Funding for open access charge: BioRange program of these Netherlands Bioinformatics Centre (NBIC); Netherlands Genomics Initiative (NGI), BSIK grant; European Commission within its FP6 Programme (to EMBRACE); under the thematic area 'Life sciences, genomics and biotechnology for health', contract number LHSG-CT-2004-512092; This work was performed within the framework of the Dutch Top Institute Pharma project D1-105, The GPCR forum.

Conflict of interest statement. None declared.

REFERENCES

1. Overington,J.P., Al-Lazikani,B. and Hopkins,A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov*, **5**, 993–996.
2. Vassilatis,D.K., Hohmann,J.G., Zeng,H., Li,F., Ranchalis,J.E., Mortrud,M.T., Brown,A., Rodriguez,S.S., Weller,J.R., Wright,A.C. *et al.* (2003) The G protein-coupled receptor repertoires of human and mouse. *Proc. Natl Acad. Sci. USA*, **100**, 4903–4908.
3. Klabunde,T. and Hessler,G. (2002) Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem*, **3**, 928–944.
4. Milligan,G. (2009) G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. *Br. J. Pharmacol.*, **158**, 5–14.
5. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
6. The Universal Protein Resource (UniProt) in 2010. (2010) *Nucleic Acids Res.*, **38**, D142–D148.
7. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
8. Stoesser,G., Baker,W., van den Broek,A., Camon,E., Garcia-Pastor,M., Kanz,C., Kulikova,T., Lombard,V., Lopez,R., Parkinson,H. *et al.* (2001) The EMBL nucleotide sequence database. *Nucleic Acids Res.*, **29**, 17–21.
9. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.
10. Kourist,R., Jochens,H., Bartsch,S., Kuipers,R., Padhi,S.K., Gall,M., Böttcher,D., Joosten,H. and Bornscheuer,U.T. (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chembiochem*, **11**, 1635–1643.
11. Kuipers,R.K., Joosten,H., van Berkel,W.J.H., Leferink,N.G.H., Rooijen,E., Ittmann,E., van Zimmeren,F., Jochens,H., Bornscheuer,U., Vriend,G. *et al.* (2010) 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins*, **78**, 2101–2113.
12. Horn,F., Bettler,E., Oliveira,L., Campagne,F., Cohen,F.E. and Vriend,G. (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.
13. Hull,D., Wolstencroft,K., Stevens,R., Goble,C., Pocock,M.R., Li,P. and Oinn,T. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.
14. Hekkelman,M.L. and Vriend,G. (2005) MRS: a fast and compact retrieval system for biological data. *Nucleic Acids Res.*, **33**, W766–W769.
15. Dowell,R.D., Jokerst,R.M., Day,A., Eddy,S.R. and Stein,L. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
16. Thornton,J. (2009) Annotations for all by all - the BioSapiens network. *Genome Biol*, **10**, 401.
17. Jimenez,R.C., Quinn,A.F., Garcia,A., Labarga,A., O'Neill,K., Martinez,F., Salazar,G.A. and Hermjakob,H. (2008) Dasty2, an Ajax protein DAS client. *Bioinformatics*, **24**, 2119–2121.
18. Brooksbank,C., Cameron,G. and Thornton,J. (2010) The European Bioinformatics Institute's data resources. *Nucleic Acids Res.*, **38**, D17–D25.
19. Okuno,Y., Tamon,A., Yabuuchi,H., Niijima,S., Minowa,Y., Tonomura,K., Kunimoto,R. and Feng,C. (2008) GLIDA: GPCR–ligand database for chemical genomics drug discovery–database and tools update. *Nucleic Acids Res.*, **36**, D907–D912.
20. Seeman,P. (1993) Drug dissociation constants for neuroreceptors and transporters. *Receptor Tables*, **Vol 2**, SZ Research, Toronto, Ontario.
21. Cutler,D. and Barbier,A. (2002) In brief. *Trends Pharmacol. Sci.*, **23**, 258–259.
22. Edvardsen,O., Reiersen,A.L., Beukers,M.W. and Kristiansen,K. (2002) tGRAP, the G-protein coupled receptors mutant database. *Nucleic Acids Res.*, **30**, 361–363.
23. Horn,F., Lau,A.L. and Cohen,F.E. (2004) Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. *Bioinformatics*, **20**, 557–568.
24. Joosten,R., Salzemann,J., Bloch,V., Stockinger,H., Berglund,A., Blanchet,C., Rudloff,E., Combet,C., Da Costa,A., Deleage,G. *et al.* (2009) {PDB_REDO: automated re-refinement of X-ray structure models in the PDB}. *J. App. Crystallogr.*, **42**, 376–384.
25. Lohse,M.J. (2010) Dimerization in GPCR mobility and signaling. *Curr. Opin. Pharmacol.*, **10**, 53–58.
26. Dean,M.K., Higgs,C., Smith,R.E., Bywater,R.P., Snell,C.R., Scott,P.D., Upton,G.J., Howe,T.J. and Reynolds,C.A. (2001) Dimerization of G-protein-coupled receptors. *J. Med. Chem.*, **44**, 4595–4614.
27. Hébert,T.E. and Bouvier,M. (1998) Structural and functional aspects of G protein-coupled receptor oligomerization. *Biochem. Cell Biol.*, **76**, 1–11.

28. Khelashvili,G., Dorff,K., Shan,J., Camacho-Artacho,M., Skrabanek,L., Vroling,B., Bouvier,M., Devi,L.A., George,S.R., Javitch,J.A. *et al.* (2010) GPCR-OKB: the G protein coupled receptor oligomer knowledge base. *Bioinformatics*, **26**, 1804–1805.
29. Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph*, **8**, 52–56, 29.
30. Krieger,E., Joo,K., Lee,J., Lee,J., Raman,S., Thompson,J., Tyka,M., Baker,D. and Karplus,K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Fouapproaches that performed well in CASP8. *Proteins*, **77(Suppl. 9)**, 114–122.
31. Oliveira,L., Paiva,A.C.M. and Vriend,G. (2002) Correlated mutation analyses on very large sequence families. *Chembiochem*, **3**, 1010–1017.
32. Oliveira,L., Paiva,A.C., Sander,C. and Vriend,G. (1994) A common step for signal transduction in G protein-coupled receptors. *Trends Pharmacol. Sci*, **15**, 170–172.
33. Oliveira,L., Paiva,P.B., Paiva,A.C.M. and Vriend,G. (2003) Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. *Proteins*, **52**, 553–560.
34. Folkertsma,S., van Noort,P., Van Durme,J., Joosten,H., Bettler,E., Fleuren,W., Oliveira,L., Horn,F., de Vlieg,J. and Vriend,G. (2004) A family-based approach reveals the function of residues in the nuclear receptor ligand-binding domain. *J. Mol. Biol*, **341**, 321–335.
35. Oliveira,L., Paiva,P.B., Paiva,A.C.M. and Vriend,G. (2003) Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins*, **52**, 544–552.
36. Ye,K., Lameijer,E.M., Beukers,M.W. and Ijzerman,A.P. (2006) A two-entropies analysis to identify functional positions in the transmembrane region of class A G protein-coupled receptors. *Proteins*, **63**, 1018–1030.
37. Oliveira,L., Paiva,A. and Vriend,G. (1993) A common motif in G-protein-coupled seven transmembrane helix receptors. *J. Comp. Aided Mol. Des.*, **7**, 649–658.
38. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
39. Goble,C.A., Bhagat,J., Aleksejevs,S., Cruickshank,D., Michaelides,D., Newman,D., Borkum,M., Bechhofer,S., Roos,M., Li,P. *et al.* (2010) myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Res.*, **38(Suppl.)**, W677–W682.
40. Harmar,A.J., Hills,R.A., Rosser,E.M., Jones,M., Buneman,O.P., Dunbar,D.R., Greenhill,S.D., Hale,V.A., Sharman,J.L., Bonner,T.I. *et al.* (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res.*, **37**, D680–D685.
41. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
42. Gloriam,D.E., Foord,S.M., Blaney,F.E. and Garland,S.L. (2009) Definition of the G protein-coupled receptor transmembrane bundle-binding pocket and calculation of receptor similarities for drug design. *J. Med. Chem*, **52**, 4429–4442.
43. Attwood,T.K., Kell,D.B., McDermott,P., Marsh,J., Pettifer,S.R. and Thorne,D. (2010) Utopia documents: linking scholarly literature with research data. *Bioinformatics*, **26**, i568–i574.
44. Braden,M.R., Parrish,J.C., Naylor,J.C. and Nichols,D.E. (2006) Molecular interaction of serotonin 5-HT2A receptor residues Phe339(6.51) and Phe340(6.52) with superpotent N-benzyl phenethylamine agonists. *Mol. Pharmacol.*, **70**, 1956–1964.